

LISTING OF CLAIMS:

1. (Previously Presented) A method for managing a textual database, comprising the steps of:

receiving textual data;

identifying a data type of the textual data;

transcribing the textual data into corresponding semantic units of words using a recognition system for the identified data type, wherein the recognition system performs transcription by decoding the textual data using a language model and phonetic dictionary of semantic units;

storing the textual data in a textual database; and

generating an index based on semantic units of words for indexing the stored textual data with the corresponding semantic units.

2. (Original) The method of claim 1, wherein the semantic units comprise syllables.

3. (Original) The method of claim 1, wherein the semantic units comprise morphemes.

4. (Original) The method of claim 1, wherein the textual data is associated with audio data, and wherein the step of indexing further comprises indexing the audio data with the semantic units.

5. (Original) The method of claim 1, wherein the step of transcribing comprises the step of time-stamping the semantic units.

6. (Canceled)

7. (Canceled)

8. (Previously Presented) The method of claim 1, wherein the step of generating an index comprises generating a hierarchical index wherein a semantic unit index points to one or more data modes.

9. (Canceled)

10. (Original) The method of claim 1, further comprising the step of converting the index into a universal index which cross-references characters of different fonts.

11. (Original) The method of claim 1, further comprising the step of searching the textual database for target textual data using the semantic unit index.

12. (Original) The method of claim 11, further comprising the step of converting a target word into a string of semantic units to perform the searching step.

13. (Original) The method of claim 12, wherein the step of converting a target word is performed automatically using a character-to-semantic unit mapping table.

14. (Original) The method of claim 11, further comprising the step of displaying search results, wherein the target textual data is displayed starting from a corresponding semantic unit in a user query and commencing one of forward and backward for a given length based on a user request.

15. (Previously Presented) A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for managing a textual database, the method comprising the steps of:

- receiving textual data;
- identifying a data type of the textual data;
- transcribing the textual data into corresponding semantic units of words using a recognition system for the identified data type, wherein the recognition system performs transcription by decoding the textual data using a language model and phonetic dictionary of semantic units;
- storing the textual data in a textual database; and
- generating an index based on semantic units of words for indexing the stored textual data with the corresponding semantic units.

16. (Previously Presented) A system for managing a textual database, comprising:

- a data type identification system for identifying a data type of textual data;
- a recognition system for transcribing the textual data into corresponding semantic units of words based on the identified data type, wherein the text recognition system performs transcription by decoding the textual data using a language model and phonetic dictionary of semantic units, wherein the recognition system comprises an OCR (optical

character recognition) system for transcribing typed text, and an AHR (automatic handwriting recognition) system for transcribing handwritten text;
a textual database for storing the textual data; and
an index generator adapted to generate an index based on semantic units of words,
wherein the textual data stored in the textual database is indexed with the corresponding semantic units.

17. (Canceled)

18. (Canceled)

19. (Original) The system of claim 16, further comprising an index converter adapted to convert the index into a universal index which cross-references characters of different fonts for a given language.

20. (Original) The system of claim 16, further comprising:
a query processor adapted to transform a search query for target textual data into corresponding semantic units; and
a search engine for searching the textual database based on the semantic units corresponding to the search query.

21. (Original) The system of claim 20, further comprising an automatic word boundary marking system that is applied to a search query.

22. (Previously Presented) The method of claim 1, wherein identifying a data type of the textual data comprises identifying data types including handwritten text and typed text.

23. (Previously Presented) The method of claim 22, wherein the different data types include handwritten text or typed text of different font or styles of a given language.

24. (Previously Presented) The method of claim 22, wherein transcribing textual data comprises decoding handwritten text using an AHR (automatic handwriting recognition) system and decoding typed text using an OCR (optical character recognition) system.

25. (Previously Presented) The method of claim 1, wherein generating an index based on semantic units of words comprises generating separate indexes for indexing the semantic units to stored textual data of different data types.

26. (Previously Presented) The method of claim 25, further comprising generating a universal index for indexing the semantic units to stored textual data of a plurality of different data types.

27. (Previously Presented) The method of claim 22, wherein generating an index based on semantic units of words comprises indexing the semantic units to stored handwritten textual data based on handwriting biometric data.

28. (Previously Presented) The method of claim 8, wherein the one or more modes of data comprises words and pictures.